

Processing of audiovisual associations in the human brain: dependency on expectations and rule complexity

Riikka Lindström^{1*}, Petri Paavilainen¹, Teija Kujala^{1,2} and Mari Tervaniemi^{1,3,4*}

¹ Cognitive Brain Research Unit, Cognitive Science, Institute of Behavioural Sciences, University of Helsinki, Helsinki, Finland

² Cicero Learning, University of Helsinki, Helsinki, Finland

³ Center of Excellence in Interdisciplinary Music Research, University of Jyväskylä, Jyväskylä, Finland

⁴ Department of Psychology, University of Jyväskylä, Jyväskylä, Finland

Edited by:

Micah M. Murray, Université de Lausanne, Switzerland

Reviewed by:

Kristina Simonyan, Mount Sinai School of Medicine, USA

Ryan A. Stevenson, Vanderbilt University Medical Center, USA

*Correspondence:

Riikka Lindström and Mari Tervaniemi, Cognitive Brain Research Unit, University of Helsinki, P.O. Box 9 (Siltavuorenpenger 1 B), FIN-00014 Helsinki, Finland.

e-mail: riikka.h.lindstrom@helsinki.fi; mari.tervaniemi@helsinki.fi

In order to respond to environmental changes appropriately, the human brain must not only be able to detect environmental changes but also to form expectations of forthcoming events. The events in the external environment often have a number of multisensory features such as pitch and form. For integrated percepts of objects and events, crossmodal processing, and crossmodally induced expectations of forthcoming events are needed. The aim of the present study was to determine whether the expectations created by visual stimuli can modulate the deviance detection in the auditory modality, as reflected by auditory event-related potentials (ERPs). Additionally, it was studied whether the complexity of the rules linking auditory and visual stimuli together affects this process. The N2 deflection of the ERP was observed in response to violations in the subjects' expectation of a forthcoming tone. Both temporal aspects and cognitive demands during the audiovisual deviance detection task modulated the brain processes involved.

Keywords: audiovisual processing, N2, mismatch negativity, event-related potential, association rule

INTRODUCTION

The processing of audiovisual associations is needed in many everyday situations like in reading and playing music scores or forming integrated percepts of multisensory objects and events. Audiovisual expectations make the environmental events more predictable and thus help guiding our actions. The human brain processes audiovisual events very rapidly: crossmodal processing effects have already been observed in the primary cortical level (see e.g., Besle et al., 2005, 2007). The N2 deflection of the auditory event-related potential (ERP) is a widely used marker of the early phases of auditory change detection. As it reflects auditory processing at cortical level, it is a feasible tool for studying how predictive visual information influences the subsequent auditory information processing. The N2 has two components: the mismatch negativity (MMN) and the N2b. MMN reflects automatic, pre-conscious detection of violations of auditory regularities (for a review, see Näätänen et al., 2007). When the auditory stimuli are attended to, the MMN is usually followed or partially overlapped by the N2b component, associated with more conscious change detection (Näätänen et al., 1982; Novak et al., 1990; Näätänen, 1992).

Both ERP and behavioral data suggest that visual cues modulate the processing of forthcoming tones in the context of music score reading (Schön and Besson, 2003, 2005; Tervaniemi et al., 2003; Yumoto et al., 2005). This is also evident for stimulus material familiar from everyday life or which are otherwise natural-like (Ullsperger et al., 2005; Aoyama et al., 2006). Ullsperger et al. (2005) found that a negative brain response with the latency of 110 ms was elicited when an auditory stimulus ("bang") was presented with an incongruent picture (e.g., a picture of a hammer

hitting a finger rather than a picture of a hammer hitting a nail). The authors proposed that this brain response could be related to the MMN change detection mechanism and, consequently, that incongruent visual stimulus could rather early and involuntarily modulate the processing of incoming auditory information. In Aoyama et al. (2006) study the auditory-visual associations were created with a piano keyboard animation. The study included both non-prediction conditions (the subjects were performing simple mental calculations while they were presented with the stimuli) and prediction conditions (the subjects were actively predicting the forthcoming tones on the basis the visual cues). Interestingly, only when the subjects actively tried to predict the forthcoming tone and when the time delay between the auditory and visual stimuli was long enough (300 ms), supratemporal brain activity was observed within 100–200 ms from the onset of the incongruent auditory stimulus.

Visually induced auditory expectations influence auditory processing also when the stimulus materials are more artificial (Widmann et al., 2004; Laine et al., 2007; see also Wu et al., 2010; Widmann et al., 2012). For example, Widmann et al. (2004), presented their non-musician subjects with visual patterns (sequences of rectangles), followed by patterns of tones ("melodies") that corresponded to the visual patterns in the majority of the trials. However, when a tone was incongruent with the visual stimulus, a negative MMN-like brain response was elicited as early as 100 ms from the onset of the tone. Laine et al. (2007), in turn, used visual cues (flashes) to predict the appearance of a deviant tone in an odd-ball paradigm. They found that the MMN was elicited not only by the deviant tones, but also by the standard tones when they were misleadingly preceded by the visual cue (i.e., when the expectation

of the forthcoming tone was violated). Recently, Wu et al. (2010) showed, using geometrical shapes and pure-tone bursts as stimulus material, that imagination can strengthen the expectation of the forthcoming tone: when subjects were presented with picture-tone pairs and their task was to actively imagine the forthcoming tone after seeing the visual cue stimulus, the discrepancy between the perceived and imagined tone elicited N2 deflection.

Thus, expectations created by visual stimuli can modulate the processing of the forthcoming tones. However, it is not yet known how the complexity of the rule linking the visual and auditory stimuli together and the timing of the stimuli affect the audiovisual deviance detection and the processing of violations in visually induced auditory expectations. Therefore, the present study investigated, by using the N2 deflection and behavioral measures, whether expectations formed by visual information modulate the processing of forthcoming tones when visual and auditory stimuli are associated with rules varying in their complexity. In half of the conditions, the associations between the tones and the circles were based on straight association rules in which each circle was paired with a particular tone. In the other conditions, the rule was more complex, with each circle being paired with a category of several tones. In addition, we used also two timing conditions: visual and auditory stimuli were presented either simultaneously or the visual stimulus was presented 300 ms before the onset of the auditory stimulus. In simultaneous stimulus presentation, there is no time for an expectation to arise. When there is a delay between the stimuli, however, the visual stimulus may evoke an expectation of the forthcoming tone. Violations of audiovisual associations in both conditions are expected to elicit the N2, but its time-course should be different, because the visually induced expectation of a forthcoming tone is hypothesized to speed up the change detection process.

MATERIALS AND METHODS

PARTICIPANTS

Twelve right handed (based on Edinburgh Handedness Inventory questionnaire Oldfield, 1971) voluntary subjects (five males, mean age 25.8 years, range 22–31 years) participated in the experiment. All subjects reported a normal or corrected-to-normal vision, normal hearing, no chronic neurological illness or mental disorder, and no reading impairment. None was a professional musician or a very actively training amateur musician, and none had received a musical education beyond conventional music lessons at school. The subjects were paid for their participation, provided their informed consent, and were treated in accordance with the Code of Ethics of the World Medical Association. The study was approved by Ethical Committee of the Department of Psychology, University of Helsinki.

STIMULI

The auditory stimuli, presented via headphones, were sinusoidal 58 dB (SPL) tones (duration 100 ms, with 10-ms rise and fall times) of different pitches. Tones were grouped into three categories: low, middle, and high tones. Each of the categories contained five tones of different pitches (see Table 1). Pitch differences between the categories were 11 semitones and within the categories one semitone.

Table 1 | Frequencies of the tones (Hz) in each category in the complex condition.

| Low tones | Middle tones | High tones |
|------------|--------------|-------------|
| 261 | 622 | 1479 |
| 277 | 659 | 1567 |
| <u>293</u> | <u>698</u> | <u>1661</u> |
| 311 | 739 | 1760 |
| 329 | 783 | 1864 |

The tones used in the simple condition are underlined.

The visual stimuli were three oval-shaped white circles presented for 50 ms in the middle of a black computer screen placed at a distance of about 180 cm from the subject. The circles were of three different diameters: large (11 cm), medium (6 cm), and small (2.6 cm), with corresponding visual angles of 3.5°, 1.9°, and 0.8°.

EXPERIMENTAL DESIGN AND PROCEDURE

The subjects were presented with circle-tone pairs in which the circles and the tones were paired together with association rules. The complexity of the rule and the inter-stimulus interval between the circle and the tone were manipulated as follows: in the simple conditions, the tones and the circles were paired with simple association rules: the large circle co-occurred with a 293-Hz tone, the medium-size circle with a 698-Hz tone, and the small circle with a 1661-Hz tone (i.e., each circle was paired with a specific tone). In the complex conditions, the large circle was randomly paired with any of the low-category tones, the medium-size circle with any of the middle-category tones, and the small circle with any of the high-category tones.

Additionally, the delay between the circle and the tone was varied as follows: in the no-delay conditions, the circle, and the tone were presented simultaneously and in the delay conditions, the circle was presented first, followed by the tone 300 ms after the onset of the circle. Thus, there were four combinations of conditions: a simple no-delay (Simp-ND) condition, a simple 300-ms delay condition (Simp-D), a complex no-delay (Comp-ND) condition, and a complex 300-ms delay condition (Comp-D). The inter-pair interval was 900 ms. In each condition, congruent ($p = 0.85$), and incongruent ($p = 0.15$) circle-tone pairs were randomly presented (for example, an incongruent circle-tone pair could be a small circle presented with a low tone).

Subjects sat in a dimly lit, electrically shielded room in front of the computer screen. The experiment started with two training sessions: first, the rules associating the circles and the tones in the simple conditions were explained to the subjects and they were presented with 66 congruent simple condition circle-tone pairs with a 300-ms delay between the stimuli. Subjects were asked to attend to the stimuli carefully and ensure that they notice the explained rules in the stimulation. Thereafter, the subjects were presented with a block of congruent ($p = 0.85$) and incongruent circle-tone pairs ($p = 0.15$) with 300 ms delay between the stimuli. The subjects were asked to press the response key always when the rule between the circle and the tone was violated. This training session continued until the subjects had correctly responded to

eight rule violations. Thereafter, a similar training session for the complex conditions followed.

After the training sessions, the experiment proper started. The subject's task was to press a response key for the incongruent circle-tone pairs. Electroencephalogram (EEG) and behavioral responses were recorded during 13 blocks: two blocks of 390 simultaneously presented simple circle-tone pairs, two blocks of 390 simultaneously presented complex circle-tone pairs, three blocks of 260 simple circle-tone pairs with 300 ms delay, and three blocks of 260 complex circle-tone pairs with 300 ms delay. Each block lasted for about 6 min. The blocks were presented in a random order. Thus, altogether 780 circle-tones pairs were presented in each of the four conditions. The subjects received no feedback of their performance. Breaks of 1–5 min separated the blocks, and the entire experiment lasted about 3 h (including breaks etc.).

DATA RECORDINGS

The EEG was continuously recorded (DC – 50 Hz; sampling rate 500 Hz) with a Neuroscan Acquire 4.3 System with a cap of 32 Ag-AgCl electrodes placed according to the International 10–20 System (Fp1, Fpz, Fp2, Fp3, Fa1, Fa2, Fp4, P5, F7, F3, Fz, F4, F8, P6, T3, C3, Cz, C4, T4, T5, P3, P4, T6, PO3, Pz, P04, and O1) and electrodes placed at the left and right mastoids. Vertical and horizontal eye movements were monitored with electrodes placed at the outer canthi of both eyes as well as above and below the right eye. The reference electrode was attached to the tip of the nose.

DATA ANALYSIS

Continuous EEG was divided into 700-ms epochs, time-locked to the onset of the auditory stimulus. The 100-ms period preceding the onset of the auditory stimulus served as the baseline. Epochs with EEG changes exceeding $\pm 100 \mu V$ at any electrode were discarded from the analysis. ERPs were averaged in each condition separately for the congruent and incongruent pairs, and were digitally filtered (band pass 1–30 Hz). Difference waves were calculated by subtracting the ERPs elicited by the incongruent pairs from the ERPs elicited by the congruent pairs. Behavioral performance was measured by defining hit percentages and response times (RTs) to hits. The RT time window for hits was 200–1200 ms from the onset of the auditory stimulus.

Mean N2 amplitudes in each condition were measured as follows: the Cz peak latency of N2 was first defined from the grand-average difference waves. Thereafter, the mean amplitudes from the individual-subject ERPs were calculated over a ± 25 -ms time period centered at this grand-average peak latency. To make the determination of the N2 peak latencies in the individual-subject data unambiguous, the data were digitally re-filtered (band pass 1–20 Hz) for the latency measurements at Cz.

The statistical significance of the N2 was analyzed with one-sample *t*-tests by comparing the mean amplitudes to zero at Cz and Pz (see Table 3). Differences in the N2 amplitudes and scalp distributions were analyzed with four-way repeated-measures analysis of variance (ANOVA). The factors were complexity of the rule (complex/simple), delay (0/300 ms), lateral electrode location (F4, C4, P4/Fz, Cz, Pz/F3, C3, P3), and anterior-posterior electrode location (F4, Fz, F3/C4, Cz, C3/P4, Pz, P3). The N2

latencies were analyzed with two-way repeated-measures ANOVA. The factors were complexity of the rule (complex/simple) and delay (0/300 ms). Hit rates and reaction times were analyzed with two-way repeated-measures ANOVAs with rule complexity (complex/simple) and delay (0/300 ms) as factors. Mauchly's Test of Sphericity served to test the assumption of equality of variances. Degrees of freedom were corrected with Greenhouse–Geisser correction when the assumption of sphericity was violated. Significant results of the ANOVA were further analyzed with Duncan multiple Range *post hoc* tests. The hit rates and reaction times were correlated to N2 mean amplitudes over fronto-central region containing electrodes F3, Fz, F4, C3, Cz, C4.

RESULTS

BEHAVIORAL RESULTS

Both the delay between the visual and auditory stimulus [$F(1,11) = 90.04$, $p < 0.001$] and the complexity of the rule [$F(1,11) = 9.19$, $p < 0.05$] had a significant main effect on the hit rate: the incongruity was more easily detected in the simple than complex conditions and in the delay than in no-delay conditions (see Table 2). RTs were shorter in the simple than complex conditions [$F(1,11) = 6.27$, $p < 0.05$] and in the delay conditions than in the no-delay conditions [$F(1,11) = 275.54$, $p < 0.001$].

ERP RESULTS

When the tone and the circle were incongruent, an N2 deflection was elicited, peaking in the delay conditions at about 250 ms and in the no-delay conditions at 500 ms from the onset of the tone (see Figures 1 and 2; Table 3). The N2 was centrally distributed, showing no clear polarity inversion at the electrode sites below the level of the Sylvian fissure.

For the N2 amplitude, an interaction between complexity and delay was found [$F(1,11) = 6.20$, $p < 0.05$]. The amplitude was significantly larger in the Simp-D condition than in the Comp-D ($p < 0.05$), Comp-ND ($p < 0.05$), and Simp-ND ($p < 0.01$) conditions. Furthermore, in the simple and complex conditions, the scalp distribution of the N2 differed significantly between the anterior and posterior parts of the scalp [$F(2,22) = 7.99$, $p < 0.05$], being smaller at the frontal electrodes than at the central ($p < 0.05$) or parietal electrodes ($p < 0.01$).

The N2 peaked earlier in the delay than in the no-delay conditions [$F(1,11) = 376.88$, $p < 0.001$]. Neither significant latency differences between the complex and simple conditions [$F(1,11) = 1.71$, $p = 0.22$] nor an interaction between the complexity and the delay [$F(1,11) = 0.455$, $p = 0.51$] were found.

Table 2 | Behavioral performance: mean RTs and hit rates in various conditions (standard deviations in parenthesis).

| Condition | RT (ms) | Hits (%) |
|-----------|----------|--------------|
| Simp-ND | 722 (71) | 68.4 (14.8) |
| Comp-ND | 738 (67) | 62.0 (14.13) |
| Simp-D | 575 (71) | 82.6 (12.2) |
| Comp-D | 589 (72) | 77.6 (12.4) |

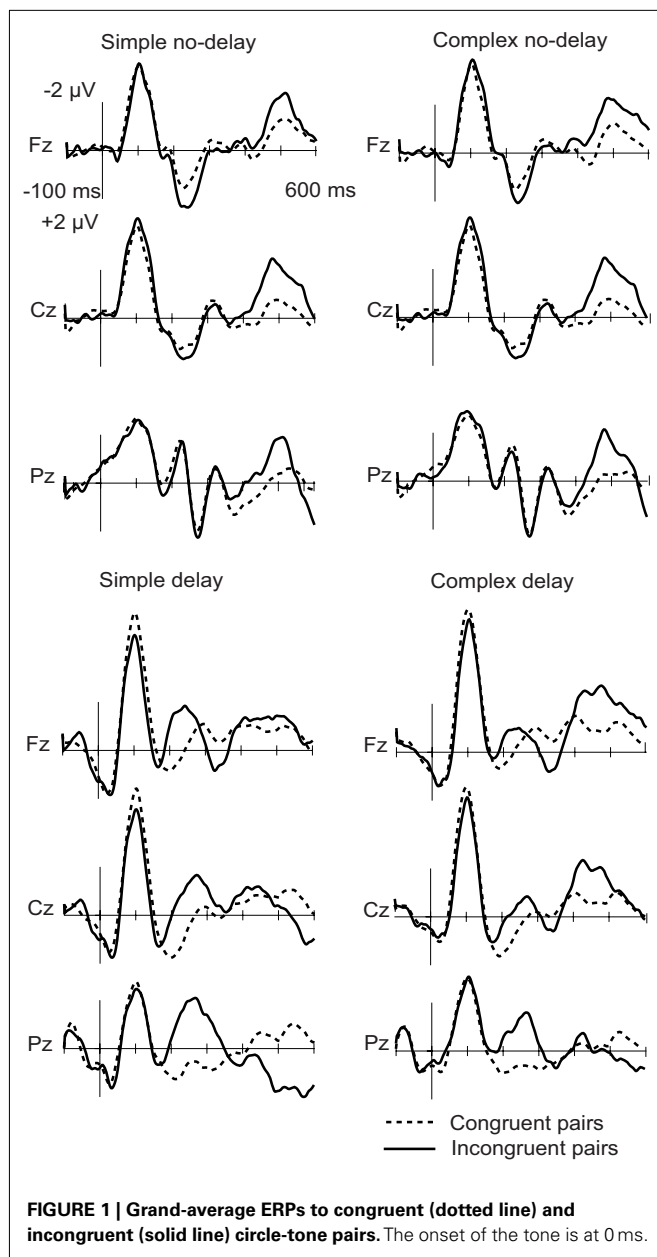


FIGURE 1 | Grand-average ERPs to congruent (dotted line) and incongruent (solid line) circle-tone pairs. The onset of the tone is at 0 ms.

CORRELATION RESULTS

There was a significant correlation between the N2 amplitude and hit rate in Comp-ND condition ($-0.78, p < 0.01$) and almost significant in simple delay condition ($-0.53, p = 0.08$). Thus, the larger the N2 amplitude, the better performance in the behavioral task was in Comp-ND condition. Also a significant correlation was found between the N2 amplitude and reaction time in complex delay condition ($0.57, p < 0.05$): the larger the N2 amplitude, the faster the reaction time was.

DISCUSSION

The aim of the present study was to determine, by varying the temporal proximity between the visual and auditory stimuli, whether the expectations created by the circles influence the processing of

the forthcoming tones. Furthermore, it was studied whether the complexity of the rules associating the visual and auditory stimuli has an effect on this process. In general, our results confirm earlier findings showing that incongruent visual information can modulate the processing of auditory information (Schön and Besson, 2003, 2005; Tervaniemi et al., 2003; Widmann et al., 2004, 2012; Ullsperger et al., 2005; Yumoto et al., 2005; Aoyama et al., 2006; Laine et al., 2007; Wu et al., 2010): ERPs elicited by the incongruent circle-tone pairs differed from those elicited by the congruent pairs in all conditions. Incongruent circle-tone pairs elicited an N2 deflection reflecting neural detection of violations of audiovisual associations. The N2 was elicited with a shorter latency when there was a 300-ms delay between the stimuli than when the stimuli were presented simultaneously, suggesting that visually induced auditory expectations speed up the processing of audiovisual associations. Hit rates correlated with the N2 amplitude in Comp-ND condition and RTs with the N2 amplitude in complex delay condition suggesting that the N2 amplitude reflected the discrimination accuracy and efficiency of complex association rules.

On the basis of some previous results (Widmann et al., 2004; Ullsperger et al., 2005; Laine et al., 2007), one could expect that the present N2 should be composed of both the MMN and N2b components. However, its topography showed no polarity inversion characteristic to MMN; see Näätänen et al., 2007). Also its latency was somewhat longer than typical for MMN. These facts suggest that the present N2 gets at least its main contribution from the N2b component (Näätänen et al., 1982; Näätänen, 1992; Novak et al., 1990; Näätänen et al., 2007; see also Widmann et al., 2012). In the present data, it is difficult to further examine this issue since we did not include an ignore condition where the possible MMN contribution could have been seen more clearly without the overlapping N2b.

The temporal proximity between the auditory and visual stimuli affected the processing of the incongruence: the RT data showed that incongruence was more rapidly detected in the delay conditions than in the no-delay conditions. Also the N2 was elicited with a shorter latency (250 vs. 500 ms) when there was a time delay between the circle and the tone (i.e., enough time for an expectation to arise) than when the stimuli were presented simultaneously. This latency difference of the N2 could be interpreted as reflecting an involvement of different brain mechanisms in the simultaneous and delayed audiovisual deviance detection. However, the scalp distribution of the N2 did not differ between these conditions. Therefore, we suggest that the brain mechanisms underlying these deviance detection processes are not fundamentally different, but the time delay facilitated the formation of expectations of the forthcoming tones, resulting in faster processing in the delay conditions. The time-course of the exogenous activity elicited by the visual and auditory stimulus naturally differs between the delay and no-delay conditions: in the no-delay conditions, the visual, and auditory activity is overlapping whereas in the delay conditions, the main part of the exogenous activity elicited by the visual stimulus probably takes place during the gap between the onset of the visual stimulus and the onset of the auditory stimulus. Consequently, in the delay conditions of the present study, it is possible that some visual activity could still continue during the baseline period (i.e., the 100-ms period preceding the onset of

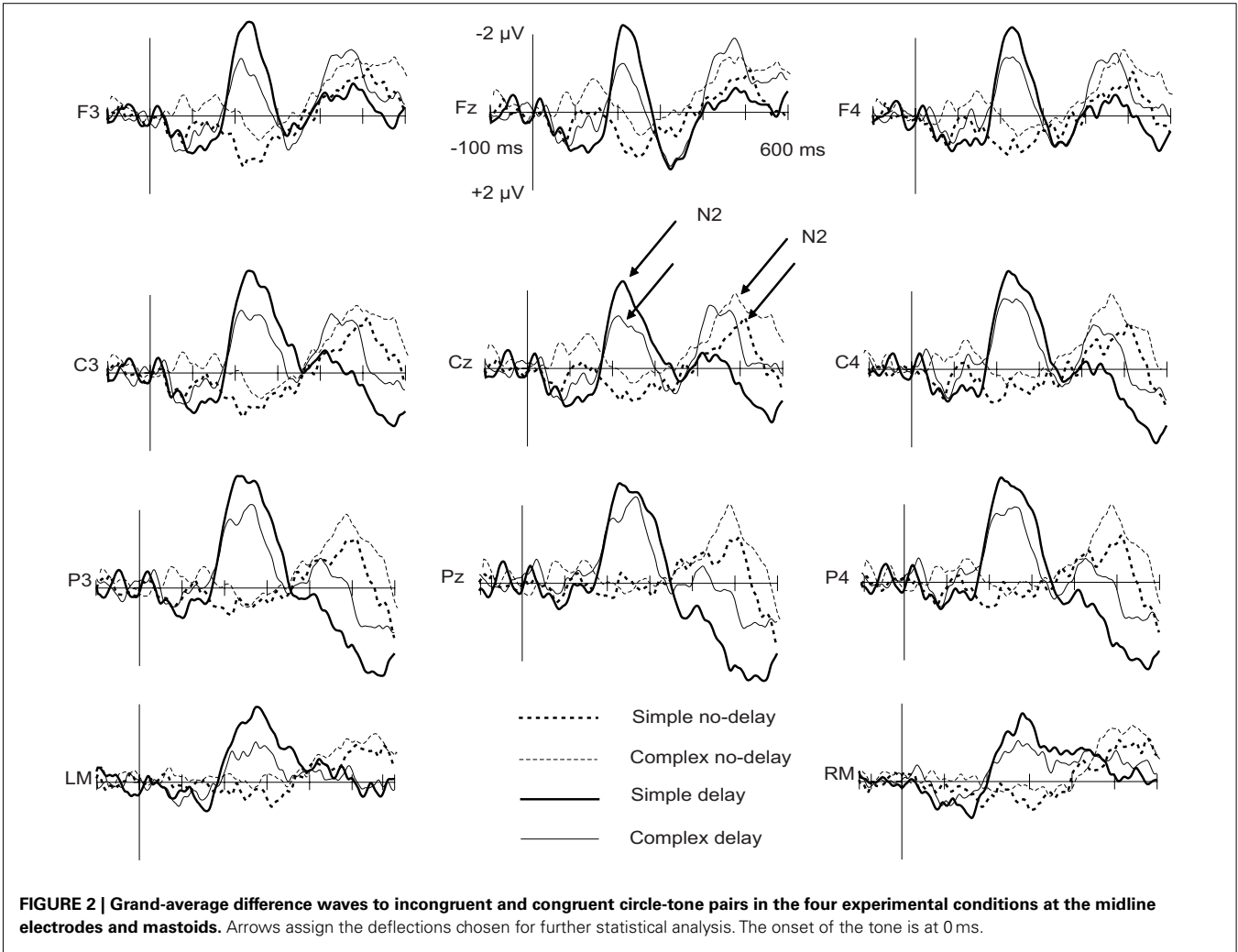


Table 3 | Mean amplitudes and latencies of the N2 deflection (standard deviations in parenthesis) and the *t*- and *p*-values.

| Condition | Electrode | Amplitude (µV) | Latency (ms) | <i>t</i> (11) | <i>p</i> |
|-----------|-----------|----------------|--------------|---------------|----------|
| Simp-ND | Cz | −1.18 (1.28) | 485 (63) | −3.19 | 0.009 |
| | Pz | −1.09 (1.27) | | −3.9 | 0.002 |
| Comp-ND | Cz | −1.56 (1.02) | 503 (42) | −5.33 | 0.000 |
| | Pz | −1.59 (1.23) | | −4.48 | 0.001 |
| Simp-D | Cz | −2.02 (1.78) | 250 (28) | −3.93 | 0.002 |
| | Pz | −2.27 (1.55) | | −5.07 | 0.000 |
| Comp-D | Cz | −1.2 (2.45) | 256 (32) | −1.7 | 0.118 |
| | Pz | −1.54 (2.01) | | −2.66 | 0.022 |

The mean amplitudes are compared to zero with one-sample *t*-tests.

the auditory stimulus) and even during the exogenous response elicited by the auditory stimulus. However, according to our view, its effects should be mainly canceled out in the incongruent minus congruent-pair difference waves.

Unlike previous studies, our experiment also addressed the impact of congruence rule complexity on the N2. We found

that the cognitive context of audiovisual associations strongly affects the visual modulation of the auditory processing: the cross-modal incongruency was more effectively detected when the rule was based on a simple association than when it was based on a more complex one. This was evident both in behavioral and N2 responses. The N2 amplitude was largest in Simp-D condition. However, the N2 amplitude differences can be (at least partially) explained by the fact that the hit percentage was highest in Simp-D condition. If the N2 is absent (or smaller in amplitude) in miss trials compared to hit trials, averaging hit and miss trials together in the ERP data (as was done in the present study) will result in larger N2 amplitudes in conditions where the hit percentage is high compared to those of lower hit percentages.

Interestingly, previous studies have shown that violations of visually induced auditory expectations can be detected by the brain as early as about 100 ms after stimulus change (Widmann et al., 2004; Aoyama et al., 2006; Laine et al., 2007). Widmann et al. (2004) observed an MMN-like negative brain response when score-like visual symbols were incongruent with the corresponding sounds. Also Laine et al. (2007) found a very early negative deflection (80–100 ms after stimulus onset) followed by the MMN to the violations of very simple audiovisual associations. Yet, the

N2 observed in the delay conditions of the present study had longer latency (250 ms) than the brain responses observed in the aforementioned studies. Importantly, the present 300-ms delay between the visual stimulus (creating the expectation) and the auditory stimulus (violating the expectation) was considerably shorter than those used by Widmann et al. (2004) and Laine et al., 2007; 1000 and 800 ms, respectively). Possibly with longer delays, the subject's expectation of the forthcoming auditory stimulus could become stronger, thus activating more efficiently the memory traces integrating auditory and visual information in the auditory cortex.

However, Aoyama et al. (2006), using the same 300-ms delay as the present study, obtained their incongruence response originating from supratemporal brain areas within 100–200 ms from the onset of the incongruent stimulus. However, they used quite natural stimulus materials (animated piano keyboards with corresponding sounds) compared to the stimulus material used in present study. Also Ullsperger et al. (2005) reported an MMN (peaking at 110 ms) to the violations of audiovisual associations familiar from everyday life (pictures and sounds about hammering). Thus, it seems that when the stimulus materials are familiar (as in Ullsperger et al., 2005) or otherwise natural (as in Aoyama et al., 2006), the processing of audiovisual deviance and visually induced auditory expectations is relatively rapid, occurring within 100–200 ms. When the stimulus materials are more artificial, like

in the present study, the processing might be prolonged and demand more conscious efforts.

To summarize, the present data indicate that predictive visual information can modulate the processing of auditory information and, further, that cognitive demands during the audiovisual deviance detection task modulate the brain processes involved. The present and previous (Widmann et al., 2004; Ullsperger et al., 2005; Aoyama et al., 2006; Laine et al., 2007) studies together suggest that the duration of the time delay between the auditory and visual stimuli and the nature of the stimulus materials together affect the processing of visually induced auditory expectations. Future research should identify how attention modulates these processes by further distinguishing the roles of automatic and controlled brain processes in audiovisual stimulus processing. When these basic mechanisms of audiovisual change detection are known, it is possible to further explore how the audiovisual associations are learned and this knowledge be effectively used in education and in rehabilitation.

ACKNOWLEDGMENTS

We would like to thank K. Reinikainen, T. Peltonen, and M. Kalske for their excellent technical support. Special acknowledgment to M. Seppänen, A. Tirhonen, H. Yli-Kaitala, and E. Kiuru for helping to generate the idea of this study. This study was supported by Academy of Finland and the University of Helsinki.

REFERENCES

- Aoyama, A., Endo, H., Honda, S., and Takeda, T. (2006). Modulation of early auditory processing by visually based sound prediction. *Brain Res.* 1068, 194–204.
- Besle, J., Cacclin, A., Mayet, R., Delpuech, C., Lecaigard, F., Giard, M. H., and Morlet, D. (2007). Audiovisual events in sensory memory. *J. Psychophysiol.* 21, 231–238.
- Besle, J., Fort, A., and Giard, M. H. (2005). Is the auditory sensory memory sensitive to visual information? *Exp. Brain Res.* 166, 337–344.
- Laine, M., Kwon, M. S., and Hämäläinen, H. (2007). Automatic auditory change detection in humans is influenced by visual-auditory associative learning. *Neuroreport* 18, 1697–1701.
- Näätänen, R. (1992). *Attention and Brain Function*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118, 2544–2590.
- Näätänen, R., Simpson, M., and Loveless, N. E. (1982). Stimulus deviance and evoked potentials. *Biol. Psychol.* 14, 53–98.
- Novak, G. P., Ritter, W., Vaughan, H. G. Jr., and Wiznitzer, M. L. (1990). Differentiation of negative event-related potentials in an auditory discrimination task. *Electroencephalogr. Clin. Neurophysiol.* 75, 255–275.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.
- Schön, D., and Besson, M. (2003). Audiovisual interactions in music reading. A reaction times and event-related potentials study. *Ann. N. Y. Acad. Sci.* 999, 193–198.
- Schön, D., and Besson, M. (2005). Visually induced auditory expectancy in music reading: a behavioral and electrophysiological study. *J. Cogn. Neurosci.* 17, 694–705.
- Tervaniemi, M., Huottilainen, M., Brattico, E., Ilmoniemi, R. J., Reinikainen, K., and Alho, K. (2003). Event-related potentials to expectancy violation in musical context. *Music Sci.* 7, 241–261.
- Ullsperger, P., Erdmann, U., Freude, G., and Dehoff, W. (2005). When sound and picture do not fit: mismatch negativity and sensory integration. *Int. J. Psychophysiol.* 59, 3–7.
- Widmann, A., Kujala, T., Tervaniemi, M., Kujala, A., and Schröger, E. (2004). From symbols to sounds: visual symbolic information activates sound representations. *Psychophysiology* 41, 709–715.
- Widmann, A., Schröger, E., Tervaniemi, M., Pakarinen, S., and Kujala, T. (2012). Mapping symbols to sounds: electrophysiological correlates of the impaired reading process in dyslexia. *Front. Psychol.* 3:60. doi:10.3389/fpsyg.2012.00060
- Wu, J., Mai, X., Yu, Z., Qin, S., and Luo, Y. (2010). Effects of discrepancy between imagined and perceived sounds on the N2 component of the event-related potential. *Psychophysiology* 47, 289–298.
- Yumoto, M., Matsuda, M., Itoh, K., Uno, A., Karino, S., Saitoh, O., Kaneko, Y., Yatomi, Y., and Kaga, K. (2005). Auditory imagery mismatch negativity elicited in musicians. *Neuroreport* 16, 1175–1178.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 17 February 2012; accepted: 03 May 2012; published online: 28 May 2012.

Citation: Lindström R, Paavilainen P, Kujala T and Tervaniemi M (2012) Processing of audiovisual associations in the human brain: dependency on expectations and rule complexity. *Front. Psychology* 3:159. doi: 10.3389/fpsyg.2012.00159 This article was submitted to *Frontiers in Auditory Cognitive Neuroscience*, a specialty of *Frontiers in Psychology*. Copyright © 2012 Lindström, Paavilainen, Kujala and Tervaniemi. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.